

Gram negative bacteria

Sinorhizobium melitoti SM; P14129
Nisseria meningitidis NM; CAM08662
Eschirichia coli EC; ABJ00328
Anaeromyxobacter A; YP_002134703
Arcobacter butzleri AB; YP_00149036
Chlamidophila pneumonia CP; Q9Z8M3
Bacteroides fragilis BF; CAH06700
Borrelia burgdorferi BB; NP_212261
Aquifex aeolicus AA; AAC07419

N-TERMINUS

SM MSATNPTRDDFAALLE
NM MSMENFAQLLE
EC MTESFAQLFE
A MENQTPRTPDVETEDFATLFAA
AB MGIEDIDLGEDFDFEQM
CP MPKQAEYTWGSKKILDNIECLTEDVAEFKDLLYTAHRITSS
BF MENLKNVAPIEDFNWDAYENGESFAGASHEELEKA
BB MRMENQKDLQEN
AA MNEFEKLLLEE

C-TERMINUS

SM EAVAQFGSSDSGASLGDILGAALKNRQNE
NM ALNSVNAAAANANAGTTS LGDLLKAKLSGEQE
EC DAIATV NKQEDANFSNMAAEAFKAAKGE
A DYREYLRRQGDGRARLGDLMEKFNRRK
AB VLKSVNDDTSMTLGDI IKDQIK
CP DQDSRTELD FKDSQGPKERKKKGGK
BF EERA EKKAASNAKSSKREETPAIQNQAASTTLGDIDALAALKEQLEGKK
BB EISSYLFKGNDEESYKPFENLLKRDE
AA KEEKVEEVTTAT TGGFKLGEILKKRWKL

DOMAIN 1

SM ESFAKTDLAE**GYVAKGIV**TAI-EKDVA**IVDV**----**GLKVEGRVPLKEFGAKAK**-DGTLKV--**GDEVEVYVERI**ENAL--**GEAVL**SREKARREES
 NM ESFTLQEMNP**GEVITA**EVVAI-DQNF**VTVNA**----**GLKSESLIDVAEFKNAQG**-EIEVKV--**GDFVTVTIES**VENGF--**GETKL**SREKAKRAAD
 EC ESLKEIETRP**GSIVRGV**VVAI-DKDV**VLVDA**----**GLKSESAIPAEQFKNAQG**-ELEIQV--**GDEV**VALDA**VEDGF**--**GETLL**SREKAKRHEA
 A SEASAGPIEE**GKVVNGTV**IQL-TKDY**AVIDI**----**GKSEGOVPISEFSTAPGGEP**AVKV--**GDKVEVLVES**RENDT--**GMVVL**SKEKADKMRI
 AB LNESFENAENNS**VVDGVIVEI**-SNDR**VLVDV**----**GQKIEGQLSVSEITIGG**--QVKYNV--**GDVIPV**MLMGNKGER--**PNISHKKVLQ**KEKFDN
 CP EEESDNEIQ**PGAILKGT**VVDI-NKDF**VVVDV**----**GLKSEGVIPMSEFIDSS**--EG-LVL--**GAEVEVYLDQ**AEDDE--**GKVVL**SREKATRORQ
 BF YDGTLNKVNDREV**VDGTVIAM**-NKRE**VVUNI**----**GKSDGIIPLNEFRYNP**--D--LKV--**GDTVEVYI**ENQEDKK--**GQLVL**SHRKARATRS
 BB YLKVLERVEL**GSRVSGTV**VUNI-MKDY**VLVDI**----**GKSEGF**IKIEEFENVP--QV-----**GDRLEAIVVR**IGGEL-**GLILSV**EKLNSLNF
 AA SIASPEEFKR**GQVVKGRV**VKV-AGDT**VFVDI**----**GKTEA**AVISKEEVPEVQ--E-----**GQIEA**VLVRFSPRIPNP**VL**SVKPLQEKKE

DOMAIN 2

SM WQRLEV**KFEAGERVEGI**IFN-QVKG**GFTVDLD**---**GAVAF**LPRSQVDIRPIR--DVTPLM--HNPQ**FEILKMD**KRR--**GNI**VVSRRTVLEESRAEQ
 NM WIALEEAMEN**GNILSGI**ING-KVKG**GGLTVMIS**---**SIRAF**LPGSLVDVRPVK--DTSHFE--**GKEIEFKVIK**LDKKR--**NNVVV**SRRVLEATLGEE
 EC WITLEKAYEDA**ETVTGVI**NG-KVKG**GFTVELN**---**GIRAF**LPGSLVDVRPVR--DTLHLE--**GKELEFKVIK**LQDKR--**NNVVV**SRRVIESENSAE
 A WDEISAACERDEL**VEGVIVG**-RVKG**GLSVDI**----**GVKAF**LPGSQVDIRPVR--NLDKLI--**GKFKFKVIK**FNKKR--**GNI**VL SRRVLEEKEREEL
 AB FVKTHGEDFED**VTIEGKI**VSVKQKG**GFI**EDAA--**GCEYF**MPMAQSYLKTQG--AI-----**GKT**VKAKVIKVNKAQ--**NSI**IVSRKKLIEESKNIK
 CP WEYILAHCEE**GSIVKQI**TR-KVKG**GGLIVDI**----**GMEAF**LPGSQIDNKKIK--NLDDYV--**GKVCEFKILK**INVER--**RNI**VVSRRELLEAERISK
 BF WDRVNAALENEE**IKGYIKC**-RTKG**GMIVDVF**---**GIEAF**LPGSQIDVKPIR--DYDVFV--**GKTMEFKVVKI**NQEF--**KNVVV**SHKALIEAELEEQ
 BB QDKVYEYIQNK**IKGKVLV**-ELPN**GYKIQ**INE--**NVSGF**MPFYLSSKSKDE--KLKR---**GSVVEFYI**LEASEAD-**GLRLIL**DRRTLEKERDLAK
 AA LEKLKEAYLNKSD**VVGKIEK**-KVKG**GFI**VDFE---**GVKAF**LPM SQAG-KLKL--E-----**GQVLT**FKVLD**FKLDS**KRPQ**IV**SHKAYLDEQREKR

DOMAIN 3

SM RSEIVQNLEE**GQVVEGV**VKNI-TDY**GAFVDLG**---**GIDGL**LHVTDMAWRRVN--HPSEILNI**GQQVKVQI**IRINQETHR**ISLGM**KQLESDF
 NM RKALLENLQ**EGSVIKGI**VKNI-TDY**GAFVDLG**---**GIDGL**LHITDLAWRRVK--HPSEVLEV**GQVEAKVLK**FDQEKQ**RVSLGM**KQLGEDP
 EC RDQLLENLQ**EGMEVKGIV**KNL-TDY**GAFVDLG**---**GVDGL**LHITDMAWKRVK--HPSEIVNV**GDEITVKVLK**FDRETR**VS**LGLKQLGEDP
 A KKETLKNLKE**GAILKGV**VKNL-TDY**GAFIDL**G---**GIDGL**LHVTDMSWGRIG--HPSEMFEV**GQEVVV**VLKFDPTSER**VS**LGLKQIQEDP
 AB DNKVAEILEKKEA**INGTI**KKI-TSY**GMFVDLG**---**GIDGL**VNYNEISYKGPV--NPANYYNE**GDEVS**VVVLAYDKAKQH**LSLSI**KAALSNP
 CP KAELIEQIS**IGEYRKGV**VKNI-TDF**GVFLDL**D---**GIDGL**LHITDMTWKRIR--HPSEMVELN**QLEVI**ILSV**DKEKGR**VALGLKQKEHNP
 BF KKEIIGKLEK**QVLEGT**VKNI-TSY**GVFIDL**G---**GVDGL**IHITDLSWGRVS--DPKEVV**ELDQKLN**VV**ILDFD**DEKKR**IALGL**KQLTPHP
 BB RIELVSSYNEED**VVDGV**VERI-TEY**GAI**VKIKN--**FVTGIL**HKRNIAFNQVE--NVEDFVR**VDKLLKI**IKINPQT**GKME**LSL**KALKANP**
 AA KQELLNLTLP**KGDVVEGK**VVKIDPN**KGVT**L**VI**EG--**VLRAF**L**PKEELS**WGRDK--NPYNYTDV**GETL**K**VKVR**VDKQ**KFIV**SLRELKGNP

DOMAIN 4

SM WDGIGAKYPVGGKISGTVTNI-TDYGAFVELEP--GIEGLIHISEMSWTKKNV-HPGKILSTSQEVVVVLEVDPTKRRISLGLKQTLNENP
NM WSGLTRRYPQDTRIFGKVSNI-TDYGAFVEIEQ--GIEGLVHVSEMDWTNKNV-HPSKVQVLDGDEVEVMILEIDEGRRRISLGMKQCQANP
EC WVAIAKRYPEGTKLTGRVTNL-TDYGCFVEIEE--GVEGLVHVSEMDWTNKNV-HPSKVNVVGDVVEVMVLDIDEERRRISLGLKQCKANP
A WHRADEKYPVGTTRVKGKVVSL-TDYGAFIELEQ--GVEGLVHVSEMSWTKRVK-HPSKLVNVGDQVEAVVLDIDPKAKRISLGMKQIEANP
AB WKEIKDQLEVGDTITVTVSNF-ESYGAFVDLGN--DIEGLLHISEISWNKNL-NPKELLTIGDEINVEVIELNVEQKRLRVSLKNLQEKP
CP WEDIEKKYPPGKRVLGKIVKL-LPYGAFIEIEE--GIEGLIHISEMSWVKNIV-DPSEVVNKGDEVEAVVLSIQKDEGKISLGLKQTERNP
BF WDALDPNLQVGDVKVGVVVM-ADYGAFIEIAP--GVEGLIHVSEMSWSQHLR-SAQDFMKVGDEVEAVVLTLDREERKMSLGIKQLKQDP
BB WDSVDVKYKIDSIVKGVVKI-LPFGAVIELDS--ELSGFLHISNFSWIRVVK-SPQELIKLQQIVEVKILEIDKENQKISLGIKQINENP
AA WEKFLKEHKEGDVIEGRVVAV-TPKGVVVEVME--GVEGFVPEEELAWEG-----KPEFKKGDLVKAKIMKIDPKRRRLILSIKRAQPRP

DOMAIN 5

SM WQAFASHPAGTEVEGEVKNK-TEFGLFIGLDG--DVDGMVHLSLDLWNRPGE-QVIEEFNKGDVVRVAVVLDVVDVKERISLGIKQLGRDAVG
NM WEEFAANHNKGDKISGAVKSI-TDFGVFVGLPG--GIDGLVHLSLDLWNTESGE-EAVRKYKKGEEVEAVVLAIDVEKERISLGIKQLEGDP
EC WQFQFAETHNKDREVGKIKSI-TDFGIFIGLDG--GIDGLVHLSLDLWNVAGE-EAVREYKKGDEIAAVVLDVDAERERISLGVKQLAEDP
A WTLLEDKYPIGTTIRGEVRNV-TDFGIFVGVVE--GIDGLVHVSDisWTERIK-HPGEKFKKGDVVEAVVLDVENERFSLGIKQAHVDP
AB FTKFVNEHKGVDVIGKGIATL-TDFGAFVNI---DVDGLLHNEEASW-ESNA-KCKSLFKKGDEVEVRIKIDKEKENISLSIKDISDSP
CP WDNIEEKYPIGLHVNAEIKNL-TNYGAFVELEP--GIEGLIHISDMSWIKKVS-HPSELFKKGNSVEAVVLSVDKESKKITLGVKQLSSNP
BF WETIEEKYPVGSKHTAKVRNF-TNFGVFVEIEE--GVDGLIHISDLSWTKKVK-HPSEFTQIGADIEVQVLEIDKENRRRLSLGHKQLEENP
BB WERLTEKYPIGRVVQGVVTNI-TKTGAFVNIIEE--GIDAYVSKFDISWLEEID--PEEYFKIGDLVNGKVLEVDKRRKRVRLGIKQLEESPTP
AA WEIFLQKHPVGSKVKGVIKI-EGSRAIVELED--GVKGVIHRSDLWTRP-K-RVEEVLKEGEEREFAVLGLEG--KYVKLGLKQLTENP

DOMAIN 6

SM EAAASGELRKNNAVSAEVIGV-NDGGIEVRLVNHEDVTAFIRRADLSRDRDEQ-RPER-FSVGQTVDARVTNFSKKDRKIQLSIKALEIAEEK
NM FGNFISVNDKGSIVKGSVKSVD-DAKGAVIALSD--EVEGYLPASEFAADRVED-LTTK-LKEGDEVEAVIVTVDRKNRSIKLSVKAKDAKESRE
EC FNNWVALNKKGAIVTGKVTA-DAKGATVELAD--GVEGYLRASEASRDRVED-ATLV-LSVGDEVEAKFTGVDRKNRAISLSVRKDEADEK
A WTTLSERHPVGARVKGRTKV-TDFGAFVEIEP--GIEGLVHVSEMKDERVEN-PRDV-VQEGQVEVKVIDMDLQDRKVALSMKAVNRDGGED
AB AKRFQDAYKLGDIVKGPVKDI-KDFGIFIKLEN--NLDGLIRNEDFGP-LKAD-EVK---NGDEIEAVIINIDTKKNRVRLSVRRLEQQQERE
CP WNEIEAMFPAGTVISGVVTKI-TAFGAFVELQN--GIEGLIHVSELS-DKPFA-KIEDIISIGENVS AKVIKLDPDHKKVSLSVKEYLADNAY
BF WDVFEFVTVGVSHEGTIEM-LDKGAVVALPY--GVEGFATPKHLVKEDGSQ-----AQMDKLEFKVIEFNKDAKRIILSHSRIFEDVAKA
BB WEDFSKSYKGDITIEVEIVEK-KSKGFQVRVYN--KIMGFISKIQLGDTKESLETFEKLNVDKLVVITSIDSKDKSVLLSYREYENQRSRE
AA WEKVKEKYKGVVVKLPVKEL-APFGAFLEFPE--GVEGLLPFSEV-----PRHIRIRKQVEYVKIIDINPKEGKITFSINALYGEKPEE

Gram positive bacteria with high G+C

Mycobacterium tuberculosis MT; CAB08883
Streptomyces coelicolor SC; CAB52054
Micrococcus luteus ML; ZP_02944046
Arthrobacter aureescens AS; YP_947812

N-TERMINUS

SC MTSSSTETTATTPQVAVNDIGNEEAFLAA
ML MTTTSSNAPQVAINDIGSAEDLLAA
AS MTITSTEKPGTPVVAINDIGTAEDFLAA

DOMAIN 1

MT IDKTIKYFNDGDI~~VE~~GTIVKVD~~RDE~~VLLDIGYKTEGVI~~PARELSIKHDVDPNEVVSV~~GDE~~VEALVLT~~KEDKEGR~~LIL~~SKKRAQYERA
SC IDETIKYFNDGDI~~V~~DGVIVKVD~~RDE~~VLLDIGYKTEGVI~~PSRELSIKHDVDPNEVVAV~~GDE~~IEALV~~LQEDKEGR~~LIL~~SKKRAQYERA
ML IDATIKYFNDGDL~~VE~~GTIVKVD~~RDE~~VLLDIGYKTEGVI~~PSRELSIKHDVDPDEVVAV~~GDT~~VEALVLT~~KEDKEGR~~LIL~~SKKRAQYERA
AS VDATIKYFNDGDL~~VE~~GTIVKVD~~RDE~~VLLDIGYKTEGVI~~PSRELSIKHDVDPGDVVAV~~GD~~QVEALVLT~~KEDKEGR~~LIL~~SKKRAQYERA

DOMAIN 2

MT WGTIEALKEKDEAVK~~GT~~VEIVVK~~GGLILD~~IGLR~~GFL~~PASLVEMRRVRDLQPYIGKEIEAKIIE~~LDKNNRV~~LSRRRAWLEQTQSEV
SC WGTIEKIKEEDGIV~~T~~GTVEIVVK~~GGLILD~~IGLR~~GFL~~PASLVEMRRVRDLQPYVGKELEAKIIE~~LDKNNRV~~LSRRRAWLEQTQSEV
ML WGDIEKIKEEDGV~~VE~~GTVEIVVK~~GGLILD~~IGLR~~GFL~~PASLVEMRRVRDLAPYIGQKLEAKIIE~~LDKNNRV~~LSRRRAWLEQTQSEV
AS WGDIEKVKEEDGV~~V~~GTVEIVVK~~GGLILD~~IGLR~~GFL~~PASLVEMRRVRDLAPYIGQ~~QIEAKIIE~~LDKNNRV~~LSRRRAWLEQTQSEV~~

DOMAIN 3

MT RSEFLNNLQKGTIRKGV~~VSSIVNF~~GAFV~~DLGGVDGL~~VHVSELSWKHIDHPSEVVQV~~GDEV~~TVEV~~LDV~~MDRER~~VSL~~SLKATQEDP
SC RQTF~~L~~TTLQKQVRS~~GV~~VSSIVNF~~GAFV~~DLGGV~~DGL~~VHVSELSWKHIDHPSEVVEV~~GQEV~~TVEV~~LDV~~MDRER~~VSL~~SLKATQEDP
ML RSNFLHKLEK~~GQVR~~NGTV~~VSSIVNF~~GAFV~~DLGGVDGL~~VHVSELSWKHIDHPSEVVEV~~GQEV~~TVEV~~LDV~~MDRER~~VSL~~SLKATQEDP
AS RSTFLNKLEK~~GQVR~~P~~GV~~VSSIVNF~~GAFV~~DLGGV~~DGL~~VHVSELSWKHIDHPSEVVEV~~GQEV~~TVEV~~LDV~~DRER~~VSL~~SLKATQEDP

DOMAIN 4

MT WRHFARTHAI~~GQIV~~PGKVTKLV~~PF~~GAFV~~R~~VEEGIEGLVHISELAERHVEVPDQVVAV~~GDDAM~~VKV~~ID~~IDLERRR~~ISLS~~SLKQANEDY
SC WQQFARTHQI~~GQV~~VP~~PF~~GKVTKLV~~PF~~GAFV~~R~~VEDEGIEGLVHISELAERHVEIPEQVVQV~~NDE~~IFVKV~~ID~~IDLERRR~~ISLS~~SLKQANEAF
ML WQLFARTHAL~~GQV~~VP~~PF~~GKVTKLV~~PF~~GAFV~~R~~VEDGIEGLVHISELASRHIDTAEQVSV~~NDE~~LFVKV~~ID~~IDLERRR~~ISLS~~SLKQANE
AS WQTFARTHAL~~GQV~~VP~~PF~~GKVTKLV~~PF~~GAFV~~R~~VEDGIEGLVHISELAVRHVELAEQVSV~~GDEL~~FKV~~ID~~IDLERRR~~ISLS~~SLKQANE

C-TERMINUS

MT TEEFDPAKYGMADSYDEQGNIIYFPEGFDAETNEWLEGFEKQRAEWEARYAEAERRHKMHTAQMEKFAAAEAAGRGADDQSSASSAP
SC GADPSTVEFDPTLYGMAASYDDQGNIIYFPEGFDPETNDWLEGYETQREAWETQYAEAQTRFEQHQAQVIKSREADEKAAAEGVDTAGAAP
ML GVDPEGTEFDPALYGMAAEYDEEGNYKYPEGFDVETNEWMEGFETERAAWEQQYADAQARWEAHKEQVRRSLQEEATAGADAGASLGGSSYSSEAP
AS GVDADSTEFDPALYGMAAEYDEEGNYKYPEGFDPESENEWLEGYETQRAAWEQQYADAQTRWEAHKKQVAQHAADDAAAATSGESDSGTTSYSSEPAAAES

Fusobacteria

Fusobacterium nucleatum FN; YP_002165428

N-TERMINUS

MEIIRAKHMGFCFGVLEAINVCNSLVEEKGRKYILGMLVHNKQVVEDMQRKGFKLVTEDELLEDMDELKEGDIVVIRAHGTSKSIHEKLLK
ERKVKVFDATCIVFNKIRQEIIEIANENGYSILFMGDKNHPEVKGIISFADDIQIFESFEEAKKIKIDIDKTYLLSTQTTLNKKKFEEIKK
YFKENYKNVVIFDKICGATAVRQKAVEDLATKVEVMIIVGDTKSSNTKKLYEISKLLNDNSYLVENEEQLDLTIFRDKKVVGITAGASTP
EETIMNIEKKIRGIYKMSNVNENQNEFLEMLEGFLPN

DOM1 QEKEFAGLTI~~G~~QTVDCVVTEVLDFGLAVDIN-~~ALKGF~~IHISEVSWKRL-DKLSDVYKVG~~DH~~IKAVVVS~~L~~DEAKRNVKLSIKKLEEDP
DOM2 WATVANEFKVDDE~~VE~~GT~~V~~TKVLPY~~GAF~~VEIKP~~GVE~~GLVHISDFSWTKKKVNVAEYVKE~~GE~~KVKVRI~~TDL~~HPEDRKLKLGIKQLVANP
DOM3 WETAEKDFAV~~G~~TI~~IK~~GKVVEVKPF~~GIF~~VEIAD~~GIDAF~~VHSSDYSWVGE---ETPKFEI~~GNE~~VELKI~~TEL~~DLNDKKIK~~GS~~LKALRKSP
DOM4 WEHAMEEYKV~~G~~TT~~VE~~KKIKTVADF~~GLF~~IELTK~~GID~~GFIP~~TQ~~FASKEFI-KNIRDKFNE~~GD~~VVKA~~Q~~VVEVNKDTQKIKLSIKKIEIEE

C-TERMINUS

EKREEREQIEKYSTSSEE

Gram positive bacteria with low G+C

<i>Clostridium perfringens</i>	firmicutes	CP; NP_563066
<i>Lactobacillus reuteri</i>	firmicutes	LR; YP_001841729
<i>Bacillus subtilis</i>	firmicutes	BS; AAC83962
<i>Lactococcus lactis</i>	firmicutes	LL; NP_266994
<i>Acholeplasma laidlawii</i>	tenericutes	AL; YP_001620864
<i>Spiroplasma kunkelii</i>	tenericutes	SK; AAP58873
<i>Mycoplasma pulmonis</i>	tenericutes	MP; NP_325961

N-TERMINUS

CP	MSENLTMSELM
LR	MAENNENKNDMLE
BS	MTEEM
LL	MNEFET

DOMAIN 1

CP	DSFELKHFHKGEIVKGVISVK-NDEIIVNIGH-FADGVVPRNEISNDKNFDIN-SINVDDIFVMVLSGDD-----GEGNVLLSKKRADAIVK
LR	ALDSIEQVKVGDVVKGEVLAIDDDRQAIVGIKDAGVEGVVPAKELSTKPVEDINDAVKVGDELDLVVISKI-GNDKEN-GSYLLSHRRLEARKV
BS	NQIDVQVPEVGDVVKGIVTKVE-DKHVDVEIINVKQSGIIPISELSSLHVEKASDVVKVDELDLKVTKVEDD-----ALILSKRAVDADRA
LL	LLNSVEDVKVRDVVKGEILSVE-NGQATVAIVGTGVEGVLTLREITNDKDADINTFVKPGDVLDLLVIKQIVGKEAEGANVYLLSLKRLEARKA
AL	MNDFIIPKKGQIEGGEVFOVK-KNYVLLDIN-AATEGTIYAEIFDRPAPEDLRKVIKKGDKVRAKVEKISE---DDRSSLIMLSRLPLLHEKN
SK	MWRIMYVKGKVFVAKVTNIT-PFGAFCELN--AAGLIHISEISDYVVDIKFVNI GDNVEVEVLDYDP-----VKKQVKLSYKNCRPELL
MP	MKVGSIKGVKVEKFN-TEKIYLNFKN-GYKGVVEKNEISDFNTKNISSFLAVNNVIDVQILEKDIKNKFLKCSFKSLRANFRKKIFS

DOMAIN 2

CP	CDDLKEAFEEEKSIKVSLEKVVKGIVGDFNGLRVFMPASQCAGRRIENLEELVGKLEVRVIEFNK-ENR-KVVVSRRVIDEEIRNNEKKAL
LR	WDDIQKKFDEGEHITAKVTQAVKGLVVDVDA-GVRGFVPASMITDHYVEDLNQFKGQELEFKIVEIEPSEN--LILSHAKVTQAVKGAEKV
BS	WEDLEKKFETKEVFEAEVKDVVKGLVVDI-GVRGFIPASLVEAHFVEDFTDYKGTLSLLVVELDRDKNR--VILSHRAVVESEQANKKQEL
LL	WTQLEGKE--GEIVTVKVTKDVKGLSVDYNGVRGFIPASMIDTYFVKDTKKFVGEIEAKIIEVNASEN--FILSRRAVVEAEAIEMRKEA
AL	MEKIQKAFDEKLEIETVVK SANDKGLVLNFGIELFLPYSLDFELKDQKDKLIGQSLVVLIEEFKADRKRPKLIATRKP IFEARRQEEQQORQEARQEE

INTERDOMAIN

LR	LVVDAGVRGFVPASMITDHYVEDLNQFKGQELEFKIVEIEPSENRLILSHKEIIQAQHEKA
----	---

DOMAIN 3

CP WSSIKEG**GE**KR**KG**KVTRLAK**F**G**A**F**V**D**I**G**G**V**E**G**L**VHLS**D**M**S**W**S**R**V**H**K**P**E**E**V**S**V**G**D**E**V**E**V**F**V**S**E**V**D**M**D**R**E**R**I**A**L**S**L**K**D**V**I**K**N**P
LR FAEL**Q**P**G**D**V**E**G**K**V**AR**M**T**N**F**G**A**F**I**D**L**G**G**V**D**G**L**V**H**V**S**E**I**S**Y**D**H**V**D**K**P**S**D**V**L**T**A**G****Q**D**V**K**V**K**V**L**S**V**D**P**E**R**E**R**I**S**L**S**I**K**Q**T**L**P**G**P
BS L**Q**S**L**E**V**G**S**V**L**D**G**K**V**Q**R**L**T**D**F**G**A**F**V**D**I**G**G**I**D**G**L**V**H**I**S**Q**L**S**H**S**H**V**E**K**P**S**D**V**V**E**G****Q**E**V**K**V**K**V**L**S**V**D**R**D**N**E**R**I**S**L**S**I**K**D**T**L**P**G**P
LL FA**Q**L**Q**E**G**D**I**V**E**G**T**V**S**R**V**T**N**F**G**A**F**V**D**L**G**G**I**D**G**L**V**H**V**S**E**L**S**H**S**R**I**K**R**P**S**D**A**V**K**P**G**D**V**V**N**V**K**I**L**K**L**D**P**E**A**G**R**L**S**L**S**L**K**A**T**Q**P**G**P**
AL L**E**T**I**T**T**G**A**V**L**E**G**V**V**E**S**F**E**T**H**A**A**F**V**R**F**E**H**V**S**G**M**L**R**I**S**Q**V**S**H**R**I**D**K**I**E**D**V**L**E**I**G**Q**V**Q**V**K**V**I**K**E**G**--**N**R**L**D**L**S**M**K**A**L**Q**P**T**P

DOMAIN 4

CP W**E**T**L**E--**G**L**K**V**G**N**V**L**G**K**V**T**N**F**I**K**V**G**A**F**V**E**V**L**P**G**I**E**G**L**V**H**I**S**E**I**T**D**E**N**I**A**K**P**S**D**I**L**E**L**G**Q**E**V**K**V**K**I**L**N**I**D**D**E**N**K**K**M**S**L**S**I**K**D**A**V**E**T**S**
LR W**D**D**I**E**E**K**A**P**V**D**S**V**L**T**G**T**V**K**R**L**T**S**F**G**A**F**V**E**V**F**P**G**V**E**G**L**V**H**I**S**Q**I**S**H**K**H**I**A**T**P**A**D**V**L**K**P**G**Q**E**V**Q**V**K**V**I**N**V**D**P**E**H**Q**R**L**G**L**S**M**K**A**L**E**E**
BS W**N**Q**I**G**E**K**V**K**P**G**D**V**L**E**G**T**V**Q**R**L**V**S**F**G**A**F**V**E**I**L**P**G**V**E**G**L**V**H**I**S**Q**I**S**N**K**H**I**G**T**P**H**E**V**L**E**E**G**Q**T**V**K**V**K**V**L**D**V**N**E**N**E**E**R**I**S**L**S**M**R**E**L**E**E**T**P**
LL W**E**Q**V**E**E**K**A**P**V**G**S**T**V**E**G**T**V**K**R**L**T**D**F**G**A**F**V**E**L**Y**P**G**V**E**G**L**V**H**V**S**Q**I**S**W**E**R**V**E**N**P**K**D**V**L**K**V**G**Q**V**V**N**V**K**V**L**D**V**K**P**A**E**E**R**I**S**L**S**I**K**A**L**E**E**A**P**
AL Y**E**A**Y**L**K**A**H**K**V**G**E**T**V**K**G**K**V**V**S**K**L**P**F**G**I**L**V**E**L**D**R**D**V**K**G**L**H**K**S**E**Y**S**W**N**P**Q**S**N**F**D**A**Y**I**K**I**D**E**I**E**A**V**I**L**S**K**D**A**K**K**E**R**I**S**L**S**K**K**V**L**E**D**N**P**

DOMAIN 5

AL W**A**K**L**N**L**R**V**G**Q**D**I**E**V**R**I**E**E**V**T**K**E**E**V**K**V**S**F**E**S**V**D**G**I**P**K**N**E**A**H**N**D**P**K**V**N**I**D**E**Y**Y**Q**V**G**D**T**V**K**A**K**V**I**E**F**N**K**Q**N**W**V**L**K**L**S**V**K**R**L**L**N**L**Q**E**R**Q**E**

C-TERMINUS

CP NE**Y**M**Q**Y**N**D**E**E**E**G**Y**S**L**A**D**L**F**K**G**L**N**L
LR R**P**K**E**D**E**N**N**N**N**G**E**N**Y**R**G**R**R**R**S**R**R**N**N**N**R**S**F**M**N**N**A**P**E**E**S**G**F**S**M**G**D**L**I**G**K**L**K**D**L**R**N**
BS K**A**D**Q**E**D**Y**R**Q**Y**Q**A**K**E**E**T**S**T**G**F**Q**L**G**D**L**I**G**D**K**L**N**K**L**K**
LL A**R**P**A**R**N**D**N**D**G**E**K**R**D**R**K**P**R**A**P**R**K**A**A**K**P**S**Y**D**L**P**E**T**Q**E**G**F**S**L**A**D**F**L**G**E**D**F**D**I**N**D**L**
AL F**E**K**Y**M**G**D**A**D**E**A**E**S**L**T**L**G**D**I**Y**T**N**L**G**K**D**K**K**K**K**
SK K**K**N**N**S**Q**I**Q**E**T**G**V**G**F**Q**M**L**S**E**K**I**N**S**L**T**G**K
MP Y**T**L**E**P**T**K**N**G**F**T**N**L**L**N**L**T**N**K**E**L**K**K**W**E**E**

Cyanobacteria

Synechococcus SY; 033698 (EC1/HS1, EC2/HS2, S1-LIKE)

N-TERMINUS

M**S**P**S**A**A**N**T**P**S**Y**D**D**F**A**L**A

DOM1 L**E**A**Q**S**L**D**S**Q**K**Q**L**V**R**G**K**V**C**E**Y**S**T**D**G**A**Y**I**D**I**G**G**K**A**P**A**F**L**P**K**R**E**A**A**L**H**A**V**L**D**L**E**A**H**L**P**K**D**E**E**L**E**F**L**V**I**R**D**Q**N**E**D**G**Q**V**T**V**S**L**R**A**L**A**L**E**Q**A**
DOM2 W**T**R**V**A**E**L**Q**E**G**Q**T**V**Q**V**K**V**T**G**S**N**K**G**G**V**T**A**D**L**E**--**G**L**R**A**F**I**P**R**S**H----**L**N**E**K**E**D**L**D**S**L**G**K**T**L**T**V**A**F**L**E**V**N**R**A**D**K**K**L**V**L**S**E**R**Q**A**A
DOM3 R**T**A**L**V**R**E**I**E**V**G**Q**L**I**N**G**K**V**T**G**L**K**P**F**G**V**F**V**D**L**G--**G**A**T**A**L**L**P**I**N**Q**I**S**Q**K**F**V**A**D**V**G**A**I**F**K**I**G**D**P**I**Q**A**L**V**V**A**I**D**N**T**K**G**R**I**S**L**S**T**K**V**L**E**N**H**P

C-TERMINUS

G**E**I**L**E**N**V**A**E**L**Q**A**S**A**A**D**R**A**E**R**A**R**K**Q**L**E**S**Q**

Deinococcus-thermus & chloroflexi

Deinococcus radiodurans DR; NP_295706
Thermus thermophilus TT; CAD30282
Chloroflexus sp. CH; YP_002570790

N-TERMINUS

DR MQHRYQARPGAFQREGFRPGGQEFMEDNTQTPAVNGGTQPTTGGDTSTQQAEGTQLDQETQSHQALNAQTEQAQTEVSOGEQARQQE
QAAKTQAQEERDYPMTMEDILAA
TT MEDKATQTPEQTFSMEEAALQET
CH MSVQDPVVTQQNEENLDWTQLL

DOMAIN 1

DR EAQEPQSVTRGDIVDGTVVFVIGQEGIAVDIGAKVEGVIPLNQLGEEPVTLEQAQEMYKPGDAIEAYVVRVDLPNSQIVLSKKRAEQDKG
TT EARLEKRVRPQQILTGVVVLVVGSEGVAVDIGAKTEGIIPFNQLTTKPLSEELRNLLSPGDEVKVVLRVDPETGQILLSRKKIEAQEK
CH DEYDYARPQRGELREGLIMHIEESGVLVSIIGTKREGIIPAQDLRQMGED---FINSLKVGDTVQVYVQEPENRDGDLILSLSMVQVAKD

DOMAIN 2

DR WRVLERMQEAEAEFEVEVLEKVRGGILVAQVEGIRAFLLPASQVDTRRVN-----DLDPYVGKPLMVKLIELNKRKRNVIIISHRAIMEAEKAKA
TT WDRIQELYEKGEFVTVTIKERVKGGVVAELDGIQGFMPASQLDLRRVP-----NLDEFVGGQVLAKIIEFHRRKGRVILSRRAVLEEEQKKA
CH WEVAEQLFHDGGITRCKVIGFNKGGLLVQFNRIIRGFVPSQVAQLHGRTAADERQOALQKMNQEIPLKVIIEVDRENRNLVLSERAATQEWKKAQ

DOMAIN 3

DR REATVGQLEPGAVFEGEVVEITDFGVFVNLGGIDGLVHRSELTYGRFNHPRDVKVQVQVIDVDEGRERINLSMKALTQDP
TT REAFLEKSLVPGQVVEGTVVEVTDGFFVNLGPDGLVHRSEITWGRFNHPREVIQKQKVKARVLSVDPEKERVNLISKALIPDP
CH KHRLLTELQPGDVLVGRVNLQTNFGAFIDLGADGLAHISELSWQRVNHPREVLQPGQEVKVVVEIDRERERIGLSIRQLQNNP

DOMAIN 4

DR WEGATERYHIGQRVKGVVNLNLTNFGAFIELEPGLEGLVHVSEMSWTKRVRHPNEVMKEGDEVEAVILRIDPRERRISLGIQRTTDDP
TT WTTVAEKYPVGTTRVRGKVVGLTQFQAFVEVEPGLEGLIHISELSWTKRPHKHPSEVVKEGDEVEAVVLRDPEERRLSLGLKQTQDPDP
CH WETIDQRYTLGQLVTPVNTVTPFGAFVQVEEAVEGLIHASELDADP-QAQRDILQPGQIVTARVISLDRQRQRMGLSLRRIGENE

DOMAIN 5

DR WSALPDRYPPGTPVKGKITGMTDFGVFMEIEEGIEGLIHISELDTQRVNNPADLFKKGDEYEAVILNIDPVEQRASLSRRRFLGGGPVP
TT WQQLTEKYPPGTVLKGKVTGVTDGFFVVEIEPGIEGLVHVSELDHKRVENPAALFKKGDEMEVVVLNIDPVEQRVLSLRKRLPPPLP

C-TERMINUS

DR QRGEQQRDYVSQGGGSRDDSTFGGGQRGGRRGGGADYAYNAKDAQGGKISTKLGDVYADLFAQFGLGNDKKDEGSSEE
TT QEEERPRRARSGERARRKGAPRRDRREYEGAVAEYNLYDASSVPTTTATVTKLGDLYGDLLASLGLLEEEAEKSRG
CH NEPAPAPADEADQPAQA

Thermotogae

Thermotoga maritima

TM; NP_229244

N-TERMINUS

MEPFEFNDEIL

DOM1 SQYEPEEFRRGQIVKGVVIGKEDDGVVVDFFGG-KSEGFVPENELIKS-----LDEYKVGENTLQILNLNYEERSILSERRPVLRKTL
DOM2 EELRKDYEEK-KPVKARIVSQTGGYNVLK--GVVSAFLPGSHSL-----RRNDPMPEKEIEVIILEMAQTR-RGPRIVVSRRALQDKK
DOM3 IEEFFSEKKVGDIVEGTVKGISNAGVEVEISE-GVRGFIPRSELSYDTRI-SPEDIVKPGQNI TAKI IELDKVK-KNVILSLKKLMPDP
DOM4 WEKVEEKYPVGVVSGEVTSIHPFGFFVRLEP-GVEGLVPRSEVFWGNARKSLEEVVSVGDLVEVEVINVDKEN-RKLTLSYRKAKGDP
DOM5 WENIEDRYNVNNVVTGKVVTGLIKQGA FVELEE-GVEGFVPVSEISWKRID-EPGEILKIGEKVKVILKIDKEN-RKITLSIKRTQENP
DOM6 WERALKELKPDSIVSGTIKKIVNSGVVVEVEEYDVEGFVPNNHLLS-----EPETGKALNLVLRIDPDEVFGGRMILSEKRYEE

C-TERMINUS

RKNIEEYKKMVEKESQKSIGDLLKKNGE